

RECEIVED

NOV 23 2001 1600

TECH CENTER 1600/2900

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/554,414B

DATE: 11/12/2001

TIME: 18:23:51

Input Set : A:\corrected Sequence listing.txt

Output Set: N:\CRF3\11122001\I554414B.raw

4 <110> APPLICANT: MCGILL UNIVERSITY  
 5 SZYF, Moshe  
 6 BHATTACHARYA, Sanjoy K.  
 7 RAMCHANDANI, Shyam  
 10 <120> TITLE OF INVENTION: DNA DEMETHYLASE, THERAPEUTIC AND  
 11 DIAGNOSTIC USES THEREOF  
 13 <130> FILE REFERENCE: 1770-183"PCT" FC/ld  
 C--> 15 <140> CURRENT APPLICATION NUMBER: US/09/554,414B  
 C--> 15 <141> CURRENT FILING DATE: 2000-09-06

15 <150> PRIOR APPLICATION NUMBER: CA 2,220,805  
 16 <151> PRIOR FILING DATE: 1997-11-12  
 18 <150> PRIOR APPLICATION NUMBER: CA 2,230,991  
 19 <151> PRIOR FILING DATE: 1998-05-11  
 21 <160> NUMBER OF SEQ ID NOS: 10  
 23 <170> SOFTWARE: FastSEQ for Windows Version 3.0  
 25 <210> SEQ ID NO: 1  
 26 <211> LENGTH: 1804  
 27 <212> TYPE: DNA  
 28 <213> ORGANISM: Unknown  
 30 <220> FEATURE:  
 31 <223> OTHER INFORMATION: cDNA encoding human demethylase  
 33 <400> SEQUENCE: 1

ENTERED

34	ccgctctg	ggcggggc	gtctccgga	ttccaagggc	tgggttacg	aagaagcgca	60
35	gagccggctg	gggagggg	tggatgcgc	cgcacccggg	gggagggcgc	tgctgcccg	120
36	agcaggagga	gggggagagc	gcggcgggcg	gcagcggcgc	tggcggcgac	tccgccatag	180
37	agcagggggg	ccagggcagc	gcgctcgctc	cgtccccggt	gagcggcg	cgcagggaag	240
38	gcgctcgggg	cggcgccgt	ggccgggggc	ggtggaagca	ggcggcccg	ggcggcgcg	300
39	tctgtggccg	tggccgtggc	cgtggccggg	gtcggggccg	tggccggggc	cggggccggg	360
40	gccgcggccg	tccccagagt	ggcggcagcg	gccttgggcg	cgacggcggc	ggcggcgcg	420
41	gcggctgcgg	cgtcggcagc	ggtggcgcg	tgcggggcg	gcgggaccc	gtccctttcc	480
42	cgtcggggag	ctcggggccg	gggcccaggg	gacccggggc	cacggagagc	gggaagagga	540
43	tggactgccc	ggccctcccc	cccggatgga	agaaggagga	agtgatccga	aatcagggc	600
44	tcagtgcctg	caagagcgat	gtctactact	tcagtccaag	tggtaagaag	ttcagaagta	660
45	aacctcagct	ggcaagatac	ctgggaaatg	ctgttgacct	tagcagtttt	gacttcagga	720
46	ccggcaagat	gatgcctagt	aaattacaga	agaacaagca	gagactccgg	aatgaccccc	780
47	tcaatcagaa	caagggtaaa	ccagacctga	acacaacatt	gccaattaga	caaactgcat	840
48	caattttcaa	gcaaccagta	accaaattca	cgaaccaccc	gagcaataag	gtgaagtcag	900
49	acccccagcg	gatgaatgaa	caaccacgtc	agcttttctg	ggagaagagg	ctacaaggac	960
50	ttagcgcatac	agatgtaaca	gaacaaatta	taaaaccat	ggagctacct	aaaggtcttc	1020
51	aaggagtcgg	tccaggtagc	aatgacgaga	cccttctgtc	tgctgtggcc	agtgttttac	1080
52	acacaagctc	tgcgccatc	acaggacaag	tctctgctgc	cgtggaaaag	aacctgctg	1140
53	tttggttaaa	cacatctcaa	cccctctgca	aagctttcat	tggtacagat	gaagacatta	1200
54	ggaaacagga	agagcgagtc	caacaagtac	gcaagaaact	ggaggaggca	ctgatggccg	1260
55	acatcctgtc	ccgggctg	gacacggagg	aagtagacat	tgacatggac	agtggagatg	1320
56	aggcgtaaga	atatgatcag	gtaactttcg	actgaccttc	cccaagagca	aattgctaga	1380
57	aacagaatta	aaacatttcc	actgggtttc	gcctgtaaga	aaaagtgtac	ctgagcacat	1440
58	agcttttttaa	tagcactaac	caatgccttt	ttagatgtat	ttttgatgta	tatatctatt	1500

## RAW SEQUENCE LISTING

DATE: 11/12/2001

PATENT APPLICATION: US/09/554,414B

TIME: 18:23:51

Input Set : A:\corrected Sequence listing.txt

Output Set: N:\CRF3\11122001\I554414B.raw

```

59 attccaaatg atgtttatgt tgaatcctag gacttaaaat gagtctttta taatagcaag      1560
60 cagggccctt cgggtgcagt gcagctttga ggccaggtgc agtctactgg aaaggtagca      1620
61 cttacgtgaa atatttggtt cccccacagt tttaataata acagatcagg agtaccaaat      1680
62 aagtttccca attaaagatt attatacttc actgtatata aacagatttt tatactttat      1740
63 tgaaagaaga tacctgtaca ttcttccatc atcactgtaa agacaaataa atgactatat      1800
64 tcac                                          1804
66 <210> SEQ ID NO: 2
67 <211> LENGTH: 411
68 <212> TYPE: PRT
69 <213> ORGANISM: Unknown
71 <220> FEATURE:
72 <223> OTHER INFORMATION: predicted amino acid of human demethylase
74 <400> SEQUENCE: 2
75 Met Arg Ala His Pro Gly Gly Gly Arg Cys Cys Pro Glu Gln Glu Glu
76 1 5 10 15
77 Gly Glu Ser Ala Ala Gly Gly Ser Gly Ala Gly Gly Asp Ser Ala Ile
78 20 25 30
79 Glu Gln Gly Gly Gln Gly Ser Ala Leu Ala Pro Ser Pro Val Ser Gly
80 35 40 45
81 Val Arg Arg Glu Gly Ala Arg Gly Gly Gly Arg Gly Arg Gly Arg Trp
82 50 55 60
83 Lys Gln Ala Gly Arg Gly Gly Gly Val Cys Gly Arg Gly Arg Gly Arg
84 65 70 75 80
85 Gly Arg Gly Arg Gly Arg Gly Arg Gly Arg Gly Arg Gly Arg Gly Arg
86 85 90 95
87 Pro Pro Ser Gly Gly Ser Gly Leu Gly Gly Asp Gly Gly Gly Cys Gly
88 100 105 110
89 Gly Gly Gly Ser Gly Gly Gly Gly Ala Pro Arg Arg Glu Pro Val Pro
90 115 120 125
91 Phe Pro Ser Gly Ser Ala Gly Pro Gly Pro Arg Gly Pro Arg Ala Thr
92 130 135 140
93 Glu Ser Gly Lys Arg Met Asp Cys Pro Ala Leu Pro Pro Gly Trp Lys
94 145 150 155 160
95 Lys Glu Glu Val Ile Arg Lys Ser Gly Leu Ser Ala Gly Lys Ser Asp
96 165 170 175
97 Val Tyr Tyr Phe Ser Pro Ser Gly Lys Lys Phe Arg Ser Lys Pro Gln
98 180 185 190
99 Leu Ala Arg Tyr Leu Gly Asn Thr Val Asp Leu Ser Ser Phe Asp Phe
100 195 200 205
101 Arg Thr Gly Lys Met Met Pro Ser Lys Leu Gln Lys Asn Lys Gln Arg
102 210 215 220
103 Leu Arg Asn Asp Pro Leu Asn Gln Asn Lys Gly Lys Pro Asp Leu Asn
104 225 230 235 240
105 Thr Thr Leu Pro Ile Arg Gln Thr Ala Ser Ile Phe Lys Gln Pro Val
106 245 250 255
107 Thr Lys Val Thr Asn His Pro Ser Asn Lys Val Lys Ser Asp Pro Gln
108 260 265 270
109 Arg Met Asn Glu Gln Pro Arg Gln Leu Phe Trp Glu Lys Arg Leu Gln
110 275 280 285

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/554,414B

DATE: 11/12/2001

TIME: 18:23:51

Input Set : A:\corrected Sequence listing.txt

Output Set: N:\CRF3\11122001\I554414B.raw

```

111 Gly Leu Ser Ala Ser Asp Val Thr Glu Gln Ile Ile Lys Thr Met Glu
112      290                      295                      300
113 Leu Pro Lys Gly Leu Gln Gly Val Gly Pro Gly Ser Asn Asp Glu Thr
114      305                      310                      315                      320
115 Leu Leu Ser Ala Val Ala Ser Ala Leu His Thr Ser Ser Ala Pro Ile
116      325                      330                      335
117 Thr Gly Gln Val Ser Ala Ala Val Glu Lys Asn Pro Ala Val Trp Leu
118      340                      345                      350
119 Asn Thr Ser Gln Pro Leu Cys Lys Ala Phe Ile Val Thr Asp Glu Asp
120      355                      360                      365
121 Ile Arg Lys Gln Glu Glu Arg Val Gln Gln Val Arg Lys Lys Leu Glu
122      370                      375                      380
123 Glu Ala Leu Met Ala Asp Ile Leu Ser Arg Ala Ala Asp Thr Glu Glu
124      385                      390                      395                      400
125 Met Asp Ile Glu Met Asp Ser Gly Asp Glu Ala
126      405                      410
128 <210> SEQ ID NO: 3
129 <211> LENGTH: 1589
130 <212> TYPE: DNA
131 <213> ORGANISM: Unknown
133 <220> FEATURE:
134 <223> OTHER INFORMATION: cDNA sequence of human dMTase2
136 <400> SEQUENCE: 3
137 cacgcgcggg cgggtgggcg gagcgccccc cctagcgggg gctgtgaagc gcggggaggg      60
138 ggccgagcgg gtggcgaagc cggcgcgcg cggctgggg gcggaggggc gaggcccgtg      120
139 ggacagaaca gctgcggcga gtggcgcgcg cgaggggagc cgaatcggcg acgagcccgg      180
140 gggtcgcaac ttgcagaagc ggcggcgcg cgcgcatcgg ccacggcggg cggaaggcc      240
141 ggggcgcaat ggagcggaag aggtgggagt gcccggcgct cccgcagggc tgggaaaggg      300
142 aagaagtgcc caggaggtcg gggctgtcgg ccggccacag ggatgtcttt tactatagcc      360
143 ccagcgggaa gaagtccgc agcaagccac aactggcacg ttacctgggc ggatccatgg      420
144 acctcagcac cttcgacttc cgcaccggaa agatgttgat gaacaagatg aataagagtc      480
145 gccagcgtgt gcgctatgat tcttccaacc aggtcaaggg caagcctgac ctgaacaccg      540
146 cgctgcctgt acggcagact gcattccatc tcaagcaacc ggtgaccaag atcaccaacc      600
147 accccagcaa caaggtcaag agcgaccgc agaaggcagt ggaccagccg aggcagcttt      660
148 tctgggagaa gaagctaagt ggattgagt cctttgacat tgcagaagaa ctggtcagga      720
149 ccattggactt gcccaagggc ctgcagggag tgggcccctg ctgtacagat gagacgctgc      780
150 tgtcagccat tgcgagtgt ctacacacca gcaccctgcc cattacaggc cagctctctg      840
151 cagccgtgga gaagaacctt ggtgtgtggc tgaacactgc acagccactg tgcaaagcct      900
152 tcatggtgac agatgacgac atcaggaagc aggaggagct ggtacagcag gtacggaagc      960
153 gacctggagga ggcactgatg gccgacatgc tagctcatgt ggaggagctt gcccgagacg      1020
154 gggaggcacc actggacaag gcctgtgcag aggaggaaga ggaggaggaa gaggaggagg      1080
155 aagagccgga gccagagcga gtgtagcaca ggtgccctgc ccaagtctgg gctgcagact      1140
156 gccttcagcc ttgcctggac caggtagggg ccagacctgt aggaggcagc cgtccacctc      1200
157 ctttccaaag cctcctgctt ccagggtctca gtgcagggag cccctgtgga ccttgaaactc      1260
158 acttgctcct gcgctgcctg gcaggaagcc ccacactgaa agcagatgag cagtgaacca      1320
159 actgagaggc cacctggaca cagtcacctc cctgcctcct tatcatagga caaggccttg      1380
160 cttggcaccg aggagctggg agccgtgttg ggtgctggag gaagtttctg gaaacacacc      1440
161 tggctatgcc caccttatgt ccctaaggct attacaggcc aggttttga ctgctccggc      1500
162 ccacagggct gccagcctc cccacactga gggtcagcag cccaccagga agtcaacttc      1560

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/554,414B

DATE: 11/12/2001

TIME: 18:23:51

Input Set : A:\corrected Sequence listing.txt

Output Set: N:\CRF3\11122001\I554414B.raw

163 cttcaataaaa ctgatggtag gaacttggtg 1589

165 <210> SEQ ID NO: 4

166 <211> LENGTH: 291

167 <212> TYPE: PRT

168 <213> ORGANISM: Unknown

170 <220> FEATURE:

171 <223> OTHER INFORMATION: predicted amino acid sequence of human dMTase2

173 <400> SEQUENCE: 4

174 Met Glu Arg Lys Arg Trp Glu Cys Pro Ala Leu Pro Gln Gly Trp Glu

175 1 5 10 15

176 Arg Glu Glu Val Pro Arg Arg Ser Gly Leu Ser Ala Gly His Arg Asp

177 20 25 30

178 Val Phe Tyr Tyr Ser Pro Ser Gly Lys Lys Phe Arg Ser Lys Pro Gln

179 35 40 45

180 Leu Ala Arg Tyr Leu Gly Gly Ser Met Asp Leu Ser Thr Phe Asp Phe

181 50 55 60

182 Arg Thr Gly Lys Met Leu Met Ser Lys Met Asn Lys Ser Arg Gln Arg

183 65 70 75 80

184 Val Arg Tyr Asp Ser Ser Asn Gln Val Lys Gly Lys Pro Asp Leu Asn

185 85 90 95

186 Thr Ala Leu Pro Val Arg Gln Thr Ala Ser Ile Phe Lys Gln Pro Val

187 100 105 110

188 Thr Lys Ile Thr Asn His Pro Ser Asn Lys Val Lys Ser Asp Pro Gln

189 115 120 125

190 Lys Ala Val Asp Gln Pro Arg Gln Leu Phe Trp Glu Lys Lys Leu Ser

191 130 135 140

192 Gly Leu Asn Ala Phe Asp Ile Ala Glu Glu Leu Val Lys Thr Met Asp

193 145 150 155 160

194 Leu Pro Lys Gly Leu Gln Gly Val Gly Pro Gly Cys Thr Asp Glu Thr

195 165 170 175

196 Leu Leu Ser Ala Ile Ala Ser Ala Leu His Thr Ser Thr Met Pro Ile

197 180 185 190

198 Thr Gly Gln Leu Ser Ala Ala Val Glu Lys Asn Pro Gly Val Trp Leu

199 195 200 205

200 Asn Thr Thr Gln Pro Leu Cys Lys Ala Phe Met Val Thr Asp Glu Asp

201 210 215 220

202 Ile Arg Lys Gln Glu Glu Leu Val Gln Gln Val Arg Lys Arg Leu Glu

203 225 230 235 240

204 Glu Ala Leu Met Ala Asp Met Leu Ala His Val Glu Glu Leu Ala Arg

205 245 250 255

206 Asp Gly Glu Ala Pro Leu Asp Lys Ala Cys Ala Glu Asp Asp Asp Glu

207 260 265 270

208 Glu Asp Glu Glu Glu Glu Glu Glu Pro Asp Pro Asp Pro Glu Met

209 275 280 285

210 Glu His Val

211 290

213 <210> SEQ ID NO: 5

214 <211> LENGTH: 1966

215 <212> TYPE: DNA

## RAW SEQUENCE LISTING

DATE: 11/12/2001

PATENT APPLICATION: US/09/554,414B

TIME: 18:23:51

Input Set : A:\corrected Sequence listing.txt

Output Set: N:\CRF3\11122001\I554414B.raw

```

216 <213> ORGANISM: Unknown
218 <220> FEATURE:
219 <223> OTHER INFORMATION: cDNA sequence of mouse dMTase1
221 <400> SEQUENCE: 5
222 gggggcggtg ccccgagaag gcggagacaa gatggccgcc catagcgctt ggaggaccta      60
223 agagggcggtg gccggggcca cgccccgggc aggagggcgc ctctgtgcgc gcccgctcta      120
224 tgatgcttgc gcgcgtcccc cgcgcgccgc gctgcgggcg gggcggtctt ccgggattcc      180
225 aagggtcggg ttacggaaga agcgcgagcg cggtcgggga gggggctgga tgcgcgcgca      240
226 cccgggggga ggccgctgct gcccgagca ggaggaggg gagagtgcgg cgggcggcag      300
227 cggcgtggc ggcgactccg ccatagagca ggggggccag ggcagcgcg tcgccccgtc      360
228 cccggtgagc ggctgcgca ggaaggcgc tcggggcggc ggccgtggcc gggggcggtg      420
229 gaagcaggcg ggccggggcg gcggcgctct tgccgtggc cggggccggg gccgtggccg      480
230 gggacgggga cggggccggg gccggggcg cgccgtccc ccgagtggcg gcagcggcct      540
231 tggcgcgac ggcggcggt gcggcgcgcg cggcagcgg ggcgcgcgcg ccccccggcg      600
232 ggagcgggtc cctttcccg cggggagcgc gggccgggg cccaggggac cccggggcac      660
233 ggagagcggg aagaggatgg attgccggc cctcccccc ggatggaaga aggaggaagt      720
234 gatccgaaaa tctgggctaa gtgctggcaa gacgatgtc tactacttca gtccaagtgg      780
235 taagaagttc agaagcaagc ctcatgtggc aaggtagctt ggaaatactg ttgatctcag      840
236 cagttttgac ttcagaactg gaaagatgat gcctagtaaa ttacagaaga acaaacagag      900
237 actgcgaaac gatcctctca atcaaaataa gggtaaacca gacttgaata caacattgcc      960
238 aattagacaa acagcatcaa ttttcaaaca accggtaacc aaagtcacaa atcatcctag     1020
239 taataaagtg aaatcagacc cacaacgaat gaatgaacag ccacgtcagc ttttctggga     1080
240 gaagaggcta caaggactta gtgcatcaga tgtaacagaa caaattataa aaaccatgga     1140
241 actacccaaq ggtcttcaag gagttggtcc aggtagcaat gatgagacct ttttatctgc     1200
242 tgttgccagt gctttgcaca caagctctgc gccaatcaca gggcaagtct ccgctgctgt     1260
243 ggaaaagaac cctgctgttt ggcttaacac atctcaacct ctctgcaaag cttttattgt     1320
244 cacagatgaa gacatcagga aacaggaaga gcgagtacag caagtacgca agaaattgga     1380
245 agaagcactg atggcagaca tcttgtcgcg agctgctgat acagaagaga tggatattga     1440
246 aatggacagt ggagtgaaag cctaagaata tgatcaggtt actttcgacc gactttcccc     1500
247 aagrgaaaat tcctagaaat tgaacaaaaa tgtttccact ggcttttgcc tgtaagaaaa     1560
248 aaaatgtacc cgagcacata gagcttttta atagcactaa ccaatgcctt tttagatgta     1620
249 tttttgatgt atatatctat tattcaaaaa atcatgttta ttttgagtcc taggacttaa     1680
250 aattagtctt ttgtaatatc aagcaggacc ctaagatgaa gctgagcttt tgatgccagg     1740
251 tgcaatctac tggaaatgta gcacttacgt aaaacatttg tttccccac agttttaata     1800
252 agaacagatc aggaattcta aataaatttc ccagttaaag attattgtga cttcactgta     1860
253 tataaacata tttttatact ttattgaaag gggacacctg tacattcttc catcatcact     1920
254 gtaaagacaa ataaatgatt atattcacia aaaaaaaaaa aaaaaa      1966
256 <210> SEQ ID NO: 6
257 <211> LENGTH: 414
258 <212> TYPE: PRT
259 <213> ORGANISM: Unknown
261 <220> FEATURE:
262 <223> OTHER INFORMATION: predicted amino acid sequence of mouse dMTase1
264 <400> SEQUENCE: 6
265 Met Arg Ala His Pro Gly Gly Gly Arg Cys Cys Pro Glu Gln Glu Glu
266 1 5 10 15
267 Gly Glu Ser Ala Ala Gly Gly Ser Gly Ala Gly Gly Asp Ser Ala Ile
268 20 25 30
269 Glu Gln Gly Gly Gln Gly Ser Ala Leu Ala Pro Ser Pro Val Ser Gly

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/554,414B

DATE: 11/12/2001

TIME: 18:23:52

Input Set : A:\corrected Sequence listing.txt

Output Set: N:\CRF3\11122001\I554414B.raw

L:15 M:270 C: Current Application Number differs, Replaced Current Application No

L:15 M:271 C: Current Filing Date differs, Replaced Current Filing Date